

IN THE UNITED STATES PATENT AND TRADEMARK OFFICEAPPLICATION PAPERS

## EXPRESS MAIL CERTIFICATE

Date 10/3/00 Label No. EL62822316845  
I hereby certify that on the day indicated above I  
deposited this paper or fee with the U.S. Postal Service  
& that it was addressed for delivery to the Commissioner  
of Patents & Trademarks, Washington, D.C. 20231 by  
"Express Mail Post Office to Addressee" service.  
PCORDES PCorden  
Name (Print) Signature

OFDANIEL JOSEPH WOLFFANDMARTIN JAMES LUCASFOR

Sub  
a1

PROVIDING BREAK POINTS IN VIRUS SCANNING OPERATION

## **BACKGROUND OF THE INVENTION**

### **Field of the Invention**

This invention relates to the field of data processing systems. More particularly, this invention relates to the field of the detection of computer viruses within computer files.

### **Description of the Prior Art**

It is known to provide anti-virus systems that are able to detect computer viruses within a computer file. A problem with such known anti-virus systems is that computer virus writers may seek to target the anti-virus system itself and exploit features of that anti-virus system in order to harm the computer system upon which the anti-virus system is running. As an example of this, it is known to produce files that are highly compressed versions of much larger files knowing that an anti-virus system will have to decompress the file in order that it can scan for viruses within it. If the decompressed file size is sufficiently large, then the amount of data requiring to be handled, even though it may contain very little information, may itself cause problems to an anti-virus system, e.g. it may exceed the amount of physical memory available requiring the extensive use of virtual memory thus significantly impacting the performance of the system conducting the anti-virus scan or in some cases even exceeding the amount of virtual memory available.

It is known to provide anti-virus scanning systems that will time-out a virus scanning operation if the system clock indicates that a predetermined duration for that virus scanning operation has been exceeded. However, such an approach has the disadvantage that a slow or stressed/overloaded (possibly deliberately) computer system may inappropriately terminate a virus scanning operation using such a simple expired time approach.

Measures that allow reliable breaks to be triggered during a virus scanning operation whilst reducing their own vulnerabilities are strongly advantageous.

## **SUMMARY OF THE INVENTION**

Viewed from one aspect the present invention provides a method of detecting computer viruses within a computer file, said method comprising the steps of:

receiving a request to scan a computer file for computer viruses;

initiating a virus scanning operation upon said computer file;

calculating during said virus scanning operation a measurement value indicative of an amount of data processing performed during said virus scanning operation;

comparing during said virus scanning said measurement value with a threshold value;

and

triggering a break in said virus operation is said measurement value exceeds said threshold value.

The invention operates by applying real time virus scanning operation metrics to the data processing being performed in order that this can be monitored and used to trigger appropriate breaks within the virus scanning operation. Using metrics associated with the amount of data processing performed provides a reliable way of resisting attacks on the anti-virus system by overloading it, whilst not exposing the system to vulnerabilities due to inappropriate breaks and possible early terminations of virus scanning operations that are not in themselves justified by an excessive amount of data processing being involved in the virus scanning operation.

Whilst the breaks triggered within the virus scanning operation could be used for various purposes, such as providing general feedback to a monitoring process, the invention is particularly useful in circumstances in which a break is used to perform a determination of whether the virus scanning operation should be terminated prior to completion.

One preferred technique for implementing the above is to monitor the size of the data processed during the data processing operation. If an excessive quantity of data is being processed during the virus scanning operation of a single computer file, then this indicates that it may be appropriate to terminate that virus scanning operation prior to its completion.

An additional degree of sophistication is provided when the size of the data processed in the virus scanning operation is compared with the size of the computer file being scanned when determining whether the amount of data being processed is excessive. It may be that a large computer file being scanned legitimately requires a large amount of data to be processed

during its scanning operation and accordingly should not be early terminated. Conversely, the type of highly compressed computer file deliberately intended to cause an overflow in the amount of data being processed would yield a much higher ratio in the amount of data being processed to the size of the computer file itself and so be distinguishable in a manner that its virus scanning may be properly early terminated.

Another possibility in obtaining a measurement value indicative of an amount of data processing being performed is to associate a complexity value with each of a plurality of tests that are applied to the computer file to check for particular computer viruses within that computer file. Some tests may be relatively quick and simple therefore having a low complexity value. Conversely, tests checking for a polymorphic virus or requiring heuristic analysis may require a much greater amount of data processing to complete and accordingly have a high complexity value. Summing the complexity values of the tests applied to a computer file and then comparing this with a threshold to trigger a break is a reliable way of regularly triggering breaks in a manner properly related to the amount of data processing being performed as discussed above.

Viewed from another aspect the present invention provides apparatus for detecting computer viruses within a computer file, said apparatus comprising:

a receiver operable to receive a request to scan a computer file for computer viruses; initiating logic operable to initiate a virus scanning operation upon said computer file; calculating logic operable to calculate during said virus scanning operation a measurement value indicative of an amount of data processing performed during said virus scanning operation;

comparing logic operable during said virus scanning to compare said measurement value with a threshold value; and

triggering logic operable to trigger a break in said virus operation is said measurement value exceeds said threshold value.

Viewed from a further aspect the invention provides a computer program product carrying a computer program for controlling a computer to detect computer viruses within a computer file, said computer program comprising:

receiver code operable to receive a request to scan a computer file for computer viruses;

initiating code operable to initiate a virus scanning operation upon said computer file;  
calculating code operable to calculate during said virus scanning operation a  
measurement value indicative of an amount of data processing performed during said virus  
scanning operation;

5 comparing code operable during said virus scanning to compare said measurement  
value with a threshold value; and

triggering code operable to trigger a break in said virus operation is said measurement  
value exceeds said threshold value.

10 The above, and other objects, features and advantages of this invention will be apparent  
from the following detailed description of illustrative embodiments which is to be read in  
connection with the accompanying drawings.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

15 Figure 1 schematically illustrates an on-access anti-virus file scanning system;

Figure 2 schematically illustrates the decompression of a computer file prior to scanning  
and the subsequent monitoring of the size of data processed;

20 Figure 3 is a flow diagram illustrating the operation of the system in accordance with  
Figure 2;

25 Figure 4 is a flow diagram illustrating an example of a determination of whether or not to  
early terminate a virus scanning operation based upon the size of data processed;

Figure 5 schematically illustrates a computer file being virus scanned and breaks  
provided within that operation based upon a sum of complexity values of applied tests;

30 Figure 6 is a flow diagram illustrating the operation of the system in accordance with  
Figure 5;

Figure 7 is a flow diagram illustrating a determination of whether or not scanning should  
be early terminated in accordance with the system of Figure 6; and

Figure 8 is a schematic representation of a general purpose computer system for performing the techniques described above;

### DESCRIPTION OF THE PREFERRED EMBODIMENTS

Figure 1 illustrates an on-access anti-virus system. A scan requesting process 2, which may be an application program interacting with a user via a display 4 and a keyboard 6, issues an access request to an operating system file system 8. This operating system file system 8, prior to servicing the access request from an associated hard disk drive 10, generates a scan request that is passed to an anti-virus system 12 together with the file concerned and further associated data. Within the anti-virus system 12, an anti-virus engine 14 working with virus definition data 16 serves to apply a plurality of tests for different known viruses and virus like behaviour to the computer file in order to detect the presence of a computer virus within that computer file. A pass or fail signal is passed back to the operating system file system 8 and used to determine whether or not the access request via the scan requesting process 2 is serviced.

Figure 2 illustrates virus scanning operation when access is made to a compressed computer file 18. In order that this compressed computer file 18 can be properly checked it is decompressed into an uncompressed file form 20 and then a sequence of tests corresponding to separate DAT driver files within the virus definition data 16 are applied to the uncompressed data. In practice the anti-virus system 12 requests a portion of the compressed file 18 to be decompressed and then applies the tests to that decompressed portion. If further portions still requiring checking, then more of the compressed file is decompressed and checked.

As illustrated in Figure 2, the different tests applied corresponding to different DAT drivers have different associated times taken for their completion. They also require differing amounts of data to be processed, e.g. differing amounts of data to be written to and read from memory or non-volatile storage. During the virus scanning operation, a tally is kept of the size of the data that has been processed so far in the virus scanning operation and when this exceeds a threshold level, a break in the virus scanning operation is triggered and a check is made as to whether or not the virus scanning operation should continue.

Figure 3 is a flow diagram illustrating the operation of Figure 2. At step 20, a scan request is received by the anti-virus system 12. At step 22, a portion of the computer file to be scanned is selected for initial processing.

At step 24, a determination is made as to whether or not the portion of data recovered from the computer file being scanned requires decompressing or unpacking prior to testing. If the data does require decompressing or unpacking, then this is performed at step 26. Step 28 updates a data process counter to take account of the decompressing or unpacking operation of step 26, and step 30 then compares this data processed counter value with a threshold value to see if it has been exceeded.

If the threshold has been exceeded, then processing proceeds to step 32 at which a determination is made as to whether or not the virus scanning operation should continue. If the virus scanning operation is not to continue, then it is terminated. If the virus scanning operation is to continue, then the data processed counter used to trigger the breaks within the virus scanning operation is reset at step 34 and processing is returned to step 36.

Step 36 selects the first DAT driver (i.e. computer virus test) to be applied to the portion of the computer file being processed. Step 38, if it determines that no decompression or unpacking is required, passes control directly to step 36.

Step 38 applies the selected test to the portion of the computer file being processed and step 40 then updates the counter of the amount of data processed in a similar manner to step 28.

Step 42 determines whether or not a threshold amount of data processed has been exceeded and if so passes processing to step 44 at which a determination is made as to whether or not to continue the virus scanning operation. If the virus scanning operation is not to continue, then the virus scanning operation is terminated. If virus scanning is to continue, then processing proceeds to step 46 at which the data processed size counter (break initiating counter) is reset and processing is returned to step 48. If the threshold value tested in step 42 was not exceeded, then step 42 passes control directly to step 48.

At step 48 a determination is made as to whether or not any more tests need to be applied to the portion of the computer file currently under test. If more tests are needed, then the next of these is selected at step 50. If no more tests are needed for that portion of the computer file, then processing proceeds to step 52 at which a determination is made as to whether or not any further portions of the computer file under test need to be scanned for computer viruses. If no further portions of the computer file under test do need to be scanned, then processing terminates. If further portions of the computer file under test do need to be subject to computer virus scanning, then processing returns to step 22 at which the next portion of the computer file for testing is selected.

Figure 4 is a flow diagram illustrating the type of processing that may be performed in steps 32 or 44 of Figure 3 in determining whether or not processing should be continued or early terminated. At step 54 a total size value for the complete amount of data processed so far in analysing the computer program under test (as compared to the amount of data that triggered the break) is updated. At step 56, a ratio of this total data processed so far compared to the file size of the computer file being scanned is calculated. The calculated ratio is compared with a threshold ratio value at step 58 and if the threshold ratio value is exceeded, then the result of the determination is to stop the scan at step 60. Conversely, if the threshold ratio is not exceeded at step 58, then step 62 sets the result of the determination to be to continue the scan operation.

Figure 5 schematically illustrates an alternative embodiment of the invention in which a complexity value scoring scheme is used to trigger breaks within the scanning operation. A computer file 64 to be virus scanned is in this case in its native form and does not require decompressing or unpacking. It will be appreciated that the complexity scoring approach could also work with compressed or packed files in providing a break triggering mechanism.

A portion of the computer file 64 to be tested is then subject to the processing associated with a series of DAT drivers within the computer virus definition data 16 of the anti-virus system 12. Each of the DAT drivers (tests) has an associated complexity value (e.g. a simple test could have a complexity value of 1 whilst a complicated heuristic test could have a complexity value of 10). The complexity values represent the amount of data processing typically required to conduct that test. A running count/tally of the total of the complexity values for the tests applied up to that point is kept and when this exceeds a threshold value a



break in the virus scanning operation is triggered and a determination made as to whether or not virus scanning operation should proceed further.

Figure 6 is a flow diagram illustrating the operation of the system of Figure 5 in which the decompression and unpacking processes have been removed. At step 66, a request to scan a computer file is received. Step 68 selects the first DAT driver to be applied to a first portion of the computer file 64. At step 70 the DAT driver selected is applied. At step 72 a complexity counter value is updated to reflect the total of the complexity values of the DAT driver tests applied up to that point. Step 74 tests whether the complexity value counter has exceeded a threshold value. If the threshold value has been exceeded, then step 76 determines whether or not the virus scanning operation should continue. If the virus scanning operation should not continue, then it is terminated. If the virus scanning operation should continue, then the break triggering counter is reset at step 78 and processing returned to step 80. If the threshold value tested at step 74 was not exceeded, then processing proceeds directly from step 74 to step 80.

Step 80 determines whether or not more DAT drivers should be applied to the portion of the computer file under test. If more DAT drivers are to be applied, then the next of these is selected at step 82 and processing is returned to step 70. If no more DAT drivers are to be applied then processing of that portion of the computer file concerned is terminated.

It will be appreciated that a further portion of the computer file may be selected for testing in accordance with the above technique as described in relation to the first example embodiment. In many practical instances, it is found that only a first portion of a computer file will in fact requiring testing.

Figure 7 illustrates an example of the processing that may be involved in the determination of step 76. At step 84, an update is made to a counter recording the total complexity of all the DAT drivers applied to the computer file under test (not just those since the last break was triggered). Step 86 then compares this total complexity value with a termination threshold value. If the termination threshold value is exceeded, then the result of the test of step 76 is set to stop by step 88. Conversely, if the threshold value is not exceeded then the determination of step 76 is set to continue by step 90.

Figure 8 schematically illustrates a general purpose computer system 92 of the type that may be used to implement the data processing described above. The general purpose computer 92 includes a central processing unit 94, a read only memory 96, a random access memory 98, a hard disk drive 100, a display driver 102 and a display 104, a user input/output unit 106 and a keyboard 108 and a mouse 110 and a network link unit 112 all linked by a common bus 114. The central processing unit 94 executes computer program instructions to provide computer code portions yielding the processing operations described above. The computer program instructions may be stored within one or more of the read only memory 96, the random access memory 98 or the hard disk drive 100. The computer program instructions may also be downloaded into the general purpose computer 92 via the network link unit 112. The computer program may be embodied as a computer program product distributed via a recording medium, such as a compact disk or a floppy disk drive, or may be downloaded from a remote source via a network link.

Although illustrative embodiments of the invention have been described in detail herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various changes and modifications can be effected therein by one skilled in the art without departing from the scope and spirit of the invention as defined by the appended claims.